

# A FEW TIPS ON WRITING PAPERS WITH MATHEMATICAL CONTENT

John N. Tsitsiklis  
MIT  
jnt@mit.edu

October 2009; this version: May 11, 2019

**Abstract.** We provide a number of suggestions on writing scientific papers with substantial mathematical content. We consider issues related to the general document structure (e.g., section organization) and style (e.g., notation), as well as issues related to seemingly minor details.

## Contents

<b>1 INTRODUCTION</b>	<b>2</b>
<b>2 BEFORE YOU START</b>	<b>2</b>
<b>3 DOCUMENT STRUCTURE</b>	<b>3</b>
3.1 Typical structure . . . . .	3
3.2 General rules . . . . .	3
3.3 The abstract . . . . .	3
3.4 The introduction . . . . .	3
3.5 The conclusions . . . . .	4
3.6 Appendices . . . . .	4
<b>4 BROAD ASPECTS OF WRITING STYLE</b>	<b>4</b>
<b>5 BROAD MATHEMATICAL STYLE</b>	<b>5</b>
<b>6 LOW-LEVEL ELEMENTS OF STYLE</b>	<b>7</b>
6.1 Low-level aspects of prose writing . . . . .	7
6.2 Low-level aspects of mathematical style . . . . .	8
6.3 Typesetting . . . . .	8
<b>7 USEFUL REFERENCES</b>	<b>9</b>

# 1 INTRODUCTION

The main goal of authoring a paper is, presumably, to convey information to the readers. Accordingly, it is in the author's interest to facilitate this information transfer. As a side benefit, as long as the technical contribution is solid, a well-written paper has a high chance of being accepted for publication subject to only minor revisions.

Different fields have somewhat different styles and conventions. These notes are likely to be most relevant to writing in the intersection of applied mathematics and engineering, e.g., systems, control, optimization, operations research, applied probability, etc., where the content is primarily mathematical, rather than experimental.

The guidelines that follow are based on a combination of our own experience and earlier writings by others on the same topic. There is no claim to originality. For example, many of the points listed below are taken directly from the sources cited at the end, especially from D. Bertsekas' "Ten Simple Rules," while others are abbreviated summaries of rules expounded in the other sources.

This document is intentionally kept short, with the hope that it will be read. We chose to highlight those issues that have come up in almost every paper that the author has written or edited.

# 2 BEFORE YOU START

1. **Audience.** Have a particular audience in mind; e.g., your fellow students, the experts, conference attendees, undergraduates, or your grandparents, and speak to them. The choice of audience determines the prior knowledge that can be assumed and the level of detail to be provided.
2. **The main point.** What is the main point (or a small number of main points) of the paper? Why should this paper exist?
3. **Make a table of contents.** In addition, for each section and subsection, write down a few bullet points about their intended contents.
4. **Collect your results.** Write down **precise** statements of the main theorems and lemmas. Be sure that you have proofs or are fully confident that you can produce them.
5. **Notation.** Record the notation that you will be using on a separate sheet. Check for **consistency, conflicts, and simplicity.**<sup>1</sup> Pick the notation that you find most convenient, and follow it. Once you have a complete draft, check the document for consistency. Sometimes you may be unsure about the most convenient notation. If so, try one particular choice, apply it consistently, and see if it works; if it does not, try again. Early agreement on notation is particularly important when multiple authors are involved.
6. **Terminology.** Pick your preferred terminology and follow it. The choice is sometimes arbitrary, but you should be consistent. For example, when talking about graphs, choose once and for all between "links," "arcs," and "edges." When talking about distributed decision making, choose between "agents," "sensors," and "nodes." The same applies to minor grammatical or spelling choices: choose between "multiagent" and "multi-agent," or between "queueing" and "queuing," or between "nonnegative" and "non-negative," and be consistent.

In real life, papers do change while being (re)written, and your choices are likely to evolve. Still, by being systematic, you will minimize errors, and you will save time in the long run.

---

<sup>1</sup>"A good notation has a subtlety and suggestiveness which at times make it almost seem like a live teacher." Attributed to Bertrand Russell, in J. R. Newman (ed.) *The World of Mathematics*, New York: Simon and Schuster, 1956.

## 3 DOCUMENT STRUCTURE

### 3.1 Typical structure

The following is a common structure.

1. Abstract
2. Introduction
3. The Model
4. Preliminaries (optional; avoid it if you can, because it increases the time until the reader gets to the core of the paper)
5. Results (usually 1-4 sections)
6. Conclusions
7. Appendices

### 3.2 General rules

**Size.** Sections should be of manageable size, with one clear theme in each section. Long sections should be broken down into subsections, and even subsubsections; ideally, you end up with **units** that are 1-3 pages long. Subsection and subsubsection headings play the role of signposts that orient the reader. So, they should be chosen carefully.

**Guidance.** Every section (or subsection) should start with some guidance on what is to happen next. (“We will now...” or “In this section,...”) A reader who does not know where a section is heading to can easily lose concentration. Similarly, at the end of a section, you can summarize what has been accomplished, and possibly mention what is coming next.

**Modules.** A typical unit (e.g., a subsection) is comprised of various items (“modules”) such as: introduction, theorems, lemmas, proofs, definitions, assumptions, explanations, figures, tables, examples, and counterexamples. Ideally, these correspond to “bullet points” that you generated when planning the document. Modules can be delineated by using subheadings or an appropriate paragraph structure.

### 3.3 The abstract

The abstract should be crisp and to the point. Do not give the history of the subject, background, motivation, or engage in marketing. Use simple, declarative sentences. A usual structure is “We consider...” followed by “We show that...” Use a minimal amount of mathematical symbols.

In a small variation to (and in violation of) the above, the abstract can also start as “Motivated by applications in ..., we ...” or “Building on the seminal work of ..., we ...”

### 3.4 The introduction

A generally undesirable but somewhat common practice is to start with background and history, before telling the reader about the main contents of the paper. However, a reader who does not know the end goal may become impatient. Instead, start by **framing the paper**, as in “In this paper we address the problem of...” This framing is similar to the abstract, but with some differences: on the one hand, it may omit the summary of the results; on the other hand, it may use some mathematical notation.

Perhaps continue with some **motivation**. Next, you can give the **background and history** of the subject, together with a review of the relevant literature. Do not just list related papers.

Put your work in context, by comparing with the closest papers in the literature. For example, “Our results generalize” (or “differ from”) “the work of . . .”.

Usually, this is also the right place for a succinct **preview** of the main results and a list of the main **contributions**. The preview and the contributions are sometimes written down as two different paragraphs, but if this becomes too repetitive, they can be interlaced within a single paragraph.

Close with a paragraph of the form “The rest of the paper is organized as follows. . .”

### 3.5 The conclusions

Many papers include a conclusions section which is just a summary of the paper, and which is rarely read. Ideally, the contents of such a summary have already been conveyed in the introduction, and there is no need for repetition. Instead, you may choose to emphasize some key ideas that emerged in the body of the paper and which would have been hard to elucidate in the introduction. Then, discuss what might lie beyond this paper, e.g., conjectures, open problems, obvious extensions, plausible extensions, alternative models, etc.

### 3.6 Appendices

Use appendices for complicated proofs that might break continuity, or for simple proofs that are somewhat straightforward but are necessary for completeness. The decision of what to relegate to an appendix may depend on the journal, page limitations, and the intended audience.

In any case, the main text should be self contained; **the reader should be able to go through the main text without ever looking at the appendix**. For example, the main text should not appeal to a “Lemma A.3,” stated and proved in the appendix. If the main text invokes this lemma, the lemma should be stated in the main text, even if the proof is in the appendix. Similarly, the main text should not use notation that is only defined in the appendix.

## 4 BROAD ASPECTS OF WRITING STYLE

This is a short compilation of hopefully useful suggestions.

1. Unless you have tremendous confidence in your writing skills, break long sentences into multiple short sentences.<sup>2</sup>
2. Choose a voice and pronoun and use it consistently. Active voice (“We will show. . .”) is preferable to passive (“It will be shown. . .”). “We” is preferable to “I”. Avoid “one,” as in “one can show that. . .”.
3. As an exception, you may switch to passive voice by writing “It can be shown that. . .”. This is a device that is sometimes employed in order to skip an elementary, perhaps pedantic, argument. The implied message is that not only “we can show. . .” but also that a competent reader should be able to reconstruct the details. If, however, the skipped argument is not elementary, a reference or a hint is in order, as in “Using a compactness argument, it can be shown that. . .,” or by providing a reference.
4. Use parallel constructions, even if your composition teacher might consider this to be a boring style. For example, the sequence of statements
  - (a) For all even integers  $n$ , property  $P_n$  holds.
  - (b) However, property  $Q_n$  holds if  $n$  is odd.should be replaced by:

---

<sup>2</sup><http://thejohnfox.com/long-sentences/>

- (a) For all even integers  $n$ , property  $P_n$  holds.
- (b) For all odd integers  $n$ , property  $Q_n$  holds.

Or even better:

- (a) If  $n$  is an even integer, then property  $P_n$  holds.
- (b) If  $n$  is an odd integer, then property  $Q_n$  holds.

## 5 BROAD MATHEMATICAL STYLE

The suggestions that follow are, in the author’s opinion, central in creating a pleasant reading experience.

1. **Proving theorems while typesetting is a dysfunctional cohabitation.** Do not try to write up a proof directly on the computer. Write all the mathematical pieces of your proof (formulas, without words) with pencil and paper. Check your proof for correctness. Once convinced, go ahead and typeset. This may feel slow, but it is faster in the long run.

In the same spirit, do not try to carry out a nontrivial revision directly on the computer. Instead mark the substance of all your changes on a printout.

2. A point worth repeating: unless you have tremendous confidence in your writing abilities, **break long sentences** into multiple sentences. This is the more so, when mathematics are involved.
3. Breaking long sentences is also useful because it imposes a **linear structure** to your argument.

For example, consider the statement: “We recall that  $y > 0$  and therefore,  $x + y > 0$ , since  $x > 0$ .” A linear version would be: “We recall that  $y > 0$ . Using also the fact  $x > 0$ , we conclude that  $x + y > 0$ .”

A more extreme example is the statement below. It is purposely nonsensical, to help you focus on the syntax, not the content.

“It follows that, since  $x > 0$ , we have  $f(x, y) > 0$ , because  $y > 0$ , and therefore (using Theorem 3.4), since  $f$  is irreducible, it is also primitive.”

A linear rearrangement could be:

“The function  $f$  is irreducible. Thus, Theorem 3.4 implies that  $f$  is primitive. Because we also have  $x > 0$  and  $y > 0$ , we conclude that  $f(x, y) > 0$ .”

**Exercise:** How would you rewrite the following sentence? “Conditioned on  $Z = 0$ , if we are further given  $X = x$ , then, if  $x$  is even,  $Y$  must be 0, whereas if  $x$  is odd,  $Y$  is equally likely to be 0 or 2 since we have assumed it is never equal to 1.”

Telltale signs of nonlinear arguments are occurrences of “then if” and “since,”

4. Make sure you have **defined every symbol** you use. Some general definitions may appear early in the paper, e.g., defining symbols such as  $\mathfrak{R}$  (the real numbers). However, most definitions should appear right before or right after the first use of a symbol. An example of the latter case (right after) is: “ $x \geq a$ , where  $a = \sqrt{\pi}$ .”
5. **Theorem or lemma statements** should be as short and **crisp** as possible. To accomplish this, define relevant terms, concepts, symbols, properties, etc., **before** the formal statement.
6. A theorem statement should not include a discussion of consequences. These should be developed outside the formal statement.

7. Introduce **mnemonic terms** or labels. A reader is much happier to read a reference to “the degeneracy condition (3.1)” rather than a reference to “Condition (3.1).”

8. **Linear proofs.** A proof can be visualized as a directed graph whose nodes are lemmas, facts, propositions, etc. (This graph would better be acyclic!) When writing down a proof, these nodes need to be laid out, in some order of your choice.

One possible order is to work backwards: “to prove A, we show that it is enough to prove B and C; here is a proof of C; now to prove B, we need to prove D and F.” This is often the way proofs are discovered. However, backwards proofs are hard to follow; a reader can easily lose track of what has been proved and what remains to be proved, lose the train of the argument, and develop doubts about the completeness of the proof.

Needless to say, a proof that mixes the forward and backward directions is comparable to medieval torture.

You should always aim at a proof with a forward structure whose graph is close to linear. This is what you need to do anyway in order to convince yourself that the proof is correct.

9. Of course, there are exceptions to the linearity rule. A common exception is to relegate proofs of intermediate results to the end of the paper (an appendix). This is fine as long as you obtain a perfectly linear proof in a thought experiment where you insert the appendix material at the right places.

10. It is often useful to **summarize the main idea** or the main flow of a proof. Such summaries are preferably given outside the proof. In any case, it should always be clear whether a particular paragraph contains motivation, explanation, handwaving, or a rigorous argument.

11. Do not use the “Voilà! There was a rabbit inside the hat!” **magic trick**, in which the magician lays out various facts, draws some consequences, and then declares “We have thus proved the following theorem.” Be upfront: have the theorem statement precede the proof.

12. Use **examples and counterexamples**.<sup>3</sup> Convince the reader that your short or long list of assumptions is there for a reason, explaining why your results fail to hold when these assumptions are removed. Alternatively, explain that the assumptions are more restrictive than necessary, for ease of exposition.

13. Use **figures** to illustrate your results and provide intuition. It is perfectly fine to use a caption of substantial length, e.g., with a self-contained graphical demonstration of a certain result.

14. Always aim at minimizing the reader’s effort. For example, avoid sudden references to terms defined many pages earlier, etc. If necessary, include **reminders** such as: “Recall the definition  $c_n = n/\log n$  [cf. Eq. (3.2)].”

15. Beware of **ambiguous quantifiers**. The phrase

“for every  $n$ , we have  $n < c$ , for some  $c$ ”

could correspond to:

“ $\forall n \exists c$ ”, i.e., “for every  $n$ , there exists some  $c$  such that  $n < c$ ”

or to

“ $\exists c \forall n$ ”, i.e., “there exists some  $c$  such that for every  $n$ , we have  $n < c$ ”

---

<sup>3</sup>“... the source of all great mathematics is the special case, the concrete example. It is frequent in mathematics that every instance of a concept of seemingly great generality is in essence the same as a small and concrete special case.” Paul R. Halmos, *I Want to be a Mathematician*, Washington: MAA Spectrum, 1985.

16. Quantifier ambiguity is often hidden inside **order of magnitude notation**. The statement  $f(n) = O(n^2)$  is unambiguous and means “there exists a constant  $c$  such that for all large enough  $n$ , we have  $f(n) \leq n^2$ .” However, a statement such as “ $T = O(n^d)$ ” is often used with different meanings, such as:

- a) there exists some  $c$  such that for all large enough  $n$  and  $d$ , we have  $T \leq cn^d$ ;
- b) for any fixed  $d$ , there exists some  $c$  such that for all  $n$  large enough, we have  $T \leq cn^d$ .

In the second version, the constant  $c$  is allowed to depend on  $d$ ; in the first, it is not.

The default interpretation is the one in (a). However, the intended meaning is sometimes the one in (b); in such cases, the intended meaning must be spelled out.

## 6 LOW-LEVEL ELEMENTS OF STYLE

The comments in this section may appear unimportant. However, when followed, they can serve to reduce distractions and improve the reading experience.

### 6.1 Low-level aspects of prose writing

1. Read every sentence to identify and remove unnecessary words or clauses.<sup>4</sup>
2. The previous rule does not apply whenever an additional word such as “then” or “that” makes parsing easier. For example:
  - (a) “If  $n$  is odd,  $n$  is nonzero” is to be replaced by “If  $n$  is odd, **then**  $n$  is nonzero.”
  - (b) “Assume  $n$  is odd” is to be replaced by “Assume **that**  $n$  is odd.” Similarly, the word “that” should usually follow the word “suppose.”

On the other hand, you may decide to omit the word “that” if it reappears shortly thereafter. For example: “Assume  $n$  is a number that divides  $m$ ” can be used in place of “Assume that  $n$  is a number that divides  $m$ .”

3. Whenever you use words such as “it,” “they,” “this,” “that,” “which,” etc. as pointers, make sure that there is no ambiguity about what these pointers are referring to.
4. Capitalize words such as “section,” “chapter,” “theorem,” etc., when they refer to a specific section/chapter/theorem. For example: “In this section, we derive a corollary of Theorem 1 in Section 2.” Notice that the first instance of “section” is not capitalized. Similarly, we write “Equation (3.1)”; or, even better, “Eq. (3.1).”
5. Note the comma at the end of the introductory clause “In this section,”. This comma is not absolutely necessary, but be consistent.
6. Note the proper use of periods and commas in the following:
  - (a) “An integer, **e.g.**, the number 3.” “E.g.” means “for example.” In American English, it is usually followed by a comma, and the same applies to “i.e.,” which means “that is.”
  - (b) “A surprising result, **cf.** [6],” where “cf.” means “compare with” and is also sometimes used to mean “see also;” note the absence of a comma here.
  - (c) “Chuck et **al.** proved Theorem 2.”

---

<sup>4</sup>As a counterexample, compare this dictum, with the following longer alternative, which essentially contradicts itself: “Most sentences in a first and preliminary draft of a paper that you write will contain words or clauses that are redundant or unnecessary — or even worse, distracting — and you should go over the paper carefully, reading and checking one sentence at a time, and work towards the elimination of such undesired features.”

7. There is an arbitrary convention of putting commas and periods inside quotes. (That is, write “this is a success,” as opposed to “this is a success”.) It is not quite logical, but is believed to be more pleasing to the eye.
8. Study the rules regarding the use of “that” and “which.” Sometimes, only one of the two is appropriate.
9. Avoid the phrases “easy,” “trivial,” “obvious,” etc.
10. Note the comma presence and absence in the following: “I will come if you go.” “If you go, I will come.”

## 6.2 Low-level aspects of mathematical style

1. The reader should be able to read sentences involving math as if they were ordinary sentences. For example, avoid “for every  $k = 1, 2, \dots, n$ ,” because it is equivalent to the ungrammatical “for every  $k$  equal to 1, 2, up to  $n$ .” You may write instead: “for every  $k \in \{1, 2, \dots\}$ ” or “for  $k = 1, 2, \dots, n$ .”
2. Do not start a sentence with a mathematical symbol.
3. Do not use  $\exists$  or  $\forall$  in the text. (However, the symbol  $\in$  is acceptable.) For example, “Therefore,  $\exists x \in S$ ” should be replaced with “Therefore, there exists some  $x \in S$ ” or “Therefore, there exists some  $x$  in  $S$ .”
4. Do not introduce unnecessary symbols. For example, replace “every prime number  $p$  is odd” with “every prime number is odd.”
5. Do not introduce new notation as a shorthand for expressions that appear only once or twice. In the same vein, introduce shorthand notation for expressions that appear more than a couple of times.
6. Punctuate sentences involving mathematical displays as if they were ordinary sentences. This rule is not universally adopted, but it is consistent with the first item above.
7. Do not say “the optimal solution” if you have not proved uniqueness. Instead, say “an optimal solution.”

## 6.3 Typesetting

The reasons for paying attention to the seemingly trivial issues of typesetting are:

- (a) better aesthetics result in a more satisfied reader;
- (b) good typesetting reduces the reader’s effort in parsing mathematical expressions;
- (c) good typesetting sends a subliminal signal to the reader that you have worked carefully on your paper, and therefore the content is more likely to be correct.

Here are some commonly occurring instances.

1. Latex’s default environment for theorems, lemmas, definitions, etc., uses italic font. Alternatives, such as putting theorems inside a box, are more effective in focusing the reader’s attention without compromising readability.
2. Newly introduced terms can be highlighted by setting them in bold or italic typeface. The common convention is to use italics, but boldface is visually easier to spot.

- Avoid inline fractions such as  $\frac{x+2}{x+3}$ , which result in small fonts and interfere with proper line spacing, unless there is a compelling reason. Instead, write  $(x+2)/(x+3)$ .
- Learn and apply the rules for breaking and aligning multiline equations. (Yes, there are such rules!)
- Insert spaces to help the reader. The displayed equation

$$f(x) < y, \quad \forall y > 0,$$

is preferable to

$$f(x) < y, \forall y > 0.$$

- Write “Eq. (2.1)” instead of “Eq. (2.1).” Can you tell the difference? It is about the space that Latex inserts after a period. The better spacing is achieved by placing the backslash character “\” and a blank space after the period.
- Replace “ $\{x|x \in R\}$ ” with “ $\{x | x \in R\}$ .” This is done using the command “\mid” or by introducing space manually, as in “\,|\,”. Similarly, replace  $\mathbf{E}[X|Y]$  with  $\mathbf{E}[X | Y]$ . This is particularly useful in cases such as

$$\mathbf{E}[X + 3 + k^2|Y = 3 + \log k + n^2].$$

The alternative,

$$\mathbf{E}[X + 3 + k^2 | Y = 3 + \log k + n^2],$$

is much easier to parse.

- The expression  $f(x + y(t + 2))$  is preferable to  $f(x + y(t + 2))$ ; the former uses “\big” outer parentheses.
- $\|x\|$  is preferable to  $||x||$ ; the former uses “\|”.
- Remember that left quotes require two keystrokes in Latex.
- Terms such as “inf,” “sup,” and “max” should be set in roman, not math, font.
- Use  $\Re$  or some other fancy form of “ $R$ ” to denote the real numbers.
- Add small horizontal spaces inside formulas, to improve readability, by using the command “\,;”; this should be done routinely in integrals, e.g., to change  $\int x dx$  to  $\int x \, dx$ .
- If you are using a complicated notation such as  $\hat{a}_{ij}^{(k)}(t)$ , you may want to define a macro. If you later decide to change it to  $\tilde{a}_{ij}^k(t)$ , you only need to redefine the macro.

## 7 USEFUL REFERENCES

The essay “How to write Mathematics,” by Paul Halmos, available at <http://www.math.washington.edu/~lind/Resources/Halmos.pdf> is a gem.

“Mathematical Writing,” by Knuth et al., available at <http://tex.loria.fr/typographie/mathwriting.pdf> is very thorough. For the impatient, the 27 rules offered in the first 6 pages are very valuable.

“Writing a Math Phase Two Paper,” by Steven Kleiman, available at <http://www.math.harvard.edu/graduate/advise/kleiman.pdf> offers plenty of useful guidance.

Finally, the present document owes a lot to Dimitri Bertsekas’ “Ten Simple Rules for Mathematical Writing,” available at [http://www.mit.edu/~dimitrib/Ten\\_Rules.pdf](http://www.mit.edu/~dimitrib/Ten_Rules.pdf).